Catalogue no. 75F0002MIE — No. 001
ISSN: 1707-2840
ISBN: 0-662-34322-0

Research Paper

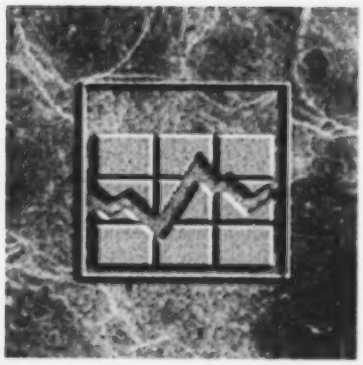
Income research paper series

User's guide for cross-sectional public-use microdata file - Survey of Labour and Income Dynamics, Reference year 1999

by Income Statistics Division

Income Statistics Division
Jean Talon Building, Ottawa, K1A 0T6

Telephone: 613 951-7355



This paper represents the views of the author and does not necessarily reflect the opinions of Statistics Canada.

Statistics Canada Statistique Canada

Canada

How to obtain more information

Specific inquiries about this product and related statistics or services should be directed to Client Services, Income Statistics Division, Statistics Canada, Ottawa, Ontario, K1A 0T6 ((613) 951-7355; (888) 297-7355; income@statcan.ca).

For information on the wide range of data available from Statistics Canada, you can contact us by calling one of our toll-free numbers. You can also contact us by e-mail or by visiting our Web site.

National inquiries line	**1 800 263-1136**
National telecommunications device for the hearing impaired	**1 800 363-7629**
Depository Services Program inquiries	**1 800 700-1033**
Fax line for Depository Services Program	**1 800 889-9734**
E-mail inquiries	**infostats@statcan.ca**
Web site	**www.statcan.ca**

Ordering and subscription information

This product, Catalogue no. 75F0002MIE2003001, is available on Internet free. Users can obtain single issues at: http://www.statcan.ca/cgi-bin/downpub/research.cgi.

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner and in the official language of their choice. To this end, the Agency has developed standards of service which its employees observe in serving its clients. To obtain a copy of these service standards, please contact Statistics Canada toll free at 1 800 263-1136.

Statistics Canada
Income Statistics Division

Income research paper series

User's guide for cross-sectional public-use microdata file - Survey of Labour and Income Dynamics, Reference year 1999

Published by authority of the Minister responsible for Statistics Canada

© Minister of Industry, 2003

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise without prior written permission from Licence Services, Marketing Division, Statistics Canada, Ottawa, Ontario, Canada K1A 0T6.

June 2003

Catalogue no. 75F0002MIE2003001

ISSN: 1707-2840

ISBN: 0-662-34322-0

Frequency: Occasional

Ottawa

La version française de cette publication est disponible sur demande (n° 75F0002MIF au catalogue).

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued cooperation and goodwill.

106-03947/

Table of Contents

1. Introduction 5

2. File structures 6

3. Using the record layouts, data dictionary and univariate distributions 6

4. Notes and definitions 9

Income definitions 9

Family definitions 14

Analytical concepts 17

Low income definitions 22

5. Guidelines for applying weights 27

6. Guidelines for release (data quality and rounding) 27

7. Confidentiality of the public-use microdata 30

8. Sources, methods and estimation procedures 31

9. SLID content. 38

10. Related products and services 41

11. Questions and comments 44

1. Introduction

The cross-sectional public-use microdata file for the Survey of Labour and Income Dynamics (SLID) is a collection of income, labour and family variables on persons in Canada and their families. SLID is an annual household survey covering the population of the 10 Canadian provinces with the exception of Indian reserves, residents of institutions and military barracks.

The Survey of Labour and Income Dynamics began collecting data for reference year 1993. Initially, SLID was designed to be, first and foremost, a longitudinal survey, with primary focus on labour and income and the relationships between them and family composition. Initially, two versions of SLID public-use microdata files were released: the first cross-sectional set covering reference year 1993 and the second longitudinal covering reference years 1993 and 1994. Both cross-sectional and longitudinal public-use files were released.

After the release of the 1993 and 1994 files, the decision was made to extend the objectives of SLID to be the primary source of cross-sectional household income data. The type of income data collected by SLID was identical to that of the former household income survey SCF (Survey of Consumer Finances), with the distinction that SLID respondents had the choice of a traditional income interview and granting permission to Statistics Canada to use their T1 income tax data.

For many years, the Survey of Consumer Finances had provided public-use microdata files (PUMFs) to meet the needs of cross-sectional household income data users. SCF PUMFs were released up to and including reference year 1997. For the purpose of standard publications, Statistics Canada has made the transition from SCF to SLID between 1995 and 1996. Therefore, SLID cross-sectional PUMFs are being made available beginning with reference year 1996. The SLID files have been designed to be analogous to those produced for the SCF.

How to cite SLID in publications

For publication of any information based on the SLID microdata files on CD-ROM (75M0010XCB), the following form of accreditation is recommended:

"This analysis is based on Statistics Canada's Survey of Labour and Income Dynamics Public Use Microdata, which contains anonymized data collected in the Survey of Labour and Income Dynamics. All computations on these microdata were prepared by (Name of user). The responsibility for the use and interpretation of these data is entirely that of the author(s)".

2. File structures

Although often referred to as one file, the SLID cross-sectional PUMF is four separate flat files: key, person, economic family and census family. To a large extent, the file structure used for SCF PUMFs has been maintained.

On the person file, there is one record per person in the sample aged 16 and over. Job characteristics such as industry, wage rates and work schedule are included on the person file and relate to the person's main job during the reference year (the job at which the most hours were worked during the year). Although SLID collects data on all jobs held during the year by each person under 70 years old, the characteristics of all other jobs are not included on the SLID PUMFs.

The person file does contain identifiers that allow a researcher to group persons into households, economic families and census families.

The sizes of the 1999 public-use files are:

Files	Number of Records	Number of Variables	Record length (bytes)
Key file	74,174	13	41
Person file	58,096	131	500
Economic family file	30,410	73	401
Census family file	33,845	70	398

3. Using the record layouts, data dictionary and univariate distributions

Additional information files are provided to assist users of the SLID public-use microdata files. For each of the four data files (key, person, economic family and census family), record layout, data dictionary and univariate distributions are provided. These information files are organized by content themes and in some cases sub-themes.

The following describes the structure of the additional information files:

A. The columns of the record layout file

- *Variable name*. This is the variable name assigned for the microdata file.
- *Data dictionary name*. This is the variable name in the internal SLID database. It is a maximum of 8 bytes long and respects the variable-naming conventions of SAS. The numerals at or near the end of the variable name refer to the source table in the master SLID database.

- *Type*. Indicates whether the variable is numeric (in the sense that it can logically be used in mathematical operations) or character.

- *Start position*. This shows the location of the variable on the public use file.

- *Format*. The format shows both the length (the number of spaces including the decimal point if there are decimal places) and the number of decimal places, if any. For example, a variable which can have values of zero (00.0) to 99.9 would have a format expressed as: 4.1. A variable which can have values of zero (00) to 99 would have a format expressed as: 2.

- *Reference year*. Reference year for the variable.

- *Long variable name*. A standardized name, with a maximum of 26 characters, which can be used to quickly identify variables, to label tables, and so on. Although still rather cryptic, it is considerably more revealing than the variable name. However, this longer name obviously excludes a lot of important information contained in the variable description shown in the data dictionary. In short, analysts are warned against making assumptions about the variable definition based on the long variable name.

- *Number of categories*. Shows the number of categories in the value set for the variable in question. Applies only to "character" variables. Numeric variables have ranges, which are specified in the data dictionary.

B. Data dictionary

The data dictionary presents the complete information about each survey variable on each of the four files. For each variable in the record layout the following information is shown: the variable name, the description or definition, code lists with descriptions or alternatively the range of values that the variable can take on, the variable type, its length (or format), and the population to which the variable pertains, i.e. for whom it is applicable.

C. Univariate distributions

These distributions are provided to allow users of the public use microdata files to verify totals that they produce. These distributions relate to the public-use files and not to the internal database; the distributions will be similar but not identical due to confidentiality processing procedures used to produce the public-use files.

For character variables, the weighted and unweighted frequencies for each code, including reserved codes (see below), are provided. For numeric variables, the values are broken into several ranges and weighted and unweighted frequencies

are provided for each range. The minimum value, the maximum value and the weighted mean (excluding reserved codes) are also provided.

Missing values and reserved codes

There are a few types of missing values on the public use file. SLID has adopted standard codes which have a particular meaning. It is important to account for reserved codes in any analysis, particularly with numeric variables. If your calculation of means or aggregates seems too high, check to ensure that you have excluded reserved codes from the calculation. With only a few exceptions, the reserved codes are the highest four values permitted according to the length of the variable. A brief explanation of reserved codes is provided below.

If the coverage of a variable does not extend to a certain population sub-group, then there are no valid values for that sub-group and the values (reserved codes) that do appear are in the form 9, 99, 9.9 and so on, which indicates that the variable is not applicable. The coverage of each variable on the file is referred to in the data dictionary as the "population".

For certain records, no valid value is available, although the value is applicable. Possibly, the respondent did not provide the information or it failed an edit in processing and the value was not imputed. Such missing values appear with a reserved code such as 7, 97, 9.7 and so on depending on the format. For certain variables, the number of missing values has been reduced through imputation. Missing values for the income variables have been entirely imputed, but most other variables may have missing values.

Finally, a few values may have been coded as 8, 98, 9.8, etc. These represent refusals to particular items in the interview. The approach for dealing with missing values of this last kind depends on the type of analysis being carried out and the extent of missing data. Although the end solution may be to exclude the records with missing values from the analysis a review should first be carried out to assess the impact of missing values on the overall representativeness of the data. Is it possible that a bias results from the missing data – for example, are the (other) characteristics of the people with missing values different from those of the observed part of the sample? It may be necessary to take into account the possible impact in some way. In all cases, analysts should note exclusions of records with missing values in their published results.

4. Notes and definitions

Income definitions

This section reviews the definitions of the main income concepts and their components. In order to highlight the relationships between them, this section is organized according to the "Classification of Income Sources", shown as a table under Total income, below.

Total income

Total income refers to income from all sources including government transfers and before deduction of federal and provincial income taxes. It may also be called income before tax (but after transfers). All sources of income are identified as belonging to either market income or government transfers.

Table A
Classification of income sources

- Total income
  - Market income
    - Earnings
      - Wages and salaries
      - Self-employment income
        - Farm
        - Non-farm
    - Investment income
    - Retirement pensions
    - Other income
  - Government transfers
    - Old Age Security and Guaranteed Income Supplement/Spouse's Allowance
    - Canada Pension Plan/Quebec Pension Plan benefits
    - Child tax benefits
    - Employment Insurance benefits
    - Workers' compensation benefits
    - GST/HST Credit
    - Provincial/territorial tax credits
    - Social assistance
    - Other government transfers

While a justification of the definition of income is not attempted here, some important inclusions and exclusions are noted.

- The concept of income covers income received while a resident of Canada or as relevant for income tax purposes in Canada. This excludes some, but not all, foreign income.

- Retirement income received as a regular pension or annuity during retirement is included, while cash withdrawals from private pension plans, including Registered Retirement Savings Plans (RRSPs), prior to retirement, are excluded.

- Realized capital gains from financial investments are excluded.

- In the CSNA and the present classification, taxes on capital gains are included in income taxes, as are taxes on RRSP withdrawals. Both capital gains (the taxable portion thereof) and RRSP withdrawals figure in the calculation of taxes, but are not part of total income in the CSNA or in this classification.

- This classification includes all refundable tax credits and benefits, including those that are not part of total income for income tax purposes, such as child tax benefits, the Goods and Services Tax Credit/Harmonized Sales Tax Credit, and other provincial or territorial tax credits. There are other smaller differences between total income here and that defined for tax purposes (see "Other income" and "Other government transfers").

- Contributions to Employment Insurance and the Canada and Quebec Pension Plans, both federal programs, are not included in income taxes, nor are they deducted from income to arrive at after-tax income. However, the Canadian System of National Accounts recently revised its definition of taxes on production to include these payroll taxes, in accordance with international recommendations on national accounting.

Market income

Market income is the sum of earnings (from employment and net self-employment), investment income, (private) retirement income, and the items under "Other income". It is equivalent to total income minus government transfers. It is also called income before taxes and transfers.

Earnings

This includes earnings from both paid employment (wages and salaries) and self-employment.

Wages and salaries

These are gross earnings from all jobs held as an employee, before payroll deductions such as income taxes, employment insurance contributions or pension plan contributions, etc. Wages and salaries include the earnings of owners of incorporated businesses, although some amounts may instead be reported as investment income. Commission income received by salespersons as well as occasional earnings for baby-sitting, for delivering papers, for cleaning, etc. are included. Overtime pay is included.

Self-employment income

This is net self-employment income, i.e. after deduction of expenses. Negative amounts (losses) are accepted. It includes income received from self-employment on own account, in partnership in an unincorporated business, or in independent professional practice. Income from roomers and boarders (excluding that received from relatives) is included. Note that because of the various inclusions, receipt of self-employment income does not necessarily mean the person held a job.

Self-employment income is subdivided into farm self-employment income and non-farm self-employment income. Farm self-employment income is reported by individuals who operate their own or a rented farm, either on own account or in partnership. Included are money receipts from the sale of farm products as well as related supplementary and assistance payments from governments. Income in kind is excluded.

Investment income

This includes interest received on bonds, deposits and savings certificates from Canadian or foreign sources, dividends received from Canadian and foreign corporate stocks, cash dividends received from insurance policies, net rental income from real estate and farms, interest received on loans and mortgages, regular income from an estate or trust fund and other investment income. Realized capital gains from the sale of assets are excluded. Negative amounts are accepted.

Retirement pensions

This is retirement pensions from all private sources, primarily employer pension plans. Amounts may be received in various forms such as annuities, superannuation or RRIFs (Registered Retirement Income Funds). Withdrawals from RRSPs (Registered Retirement Savings Plans) are not included in retirement pensions. However, they are taken into account as necessary for the estimation of certain government transfers and taxes. For data obtained from administrative records, income withdrawn from RRSPs before the age of 65 is treated as RRSP withdrawals, and income withdrawn from RRSPs at ages 65 or older is treated as retirement pensions. Retirement pensions may also be called pension income.

Government transfers

Government transfers include all direct payments from federal, provincial and municipal governments to individuals or families. See the table "Classification of Income Sources" for a list of the government transfers identified separately in the latest reference year. It should be noted that many features of the tax system also carry out social policy functions but are not government transfers per se. The tax system uses deductions and non-refundable tax credits, for example, to reduce the amount of tax payable, without providing a direct income.

Child tax benefits

Federal child tax benefits began in 1993 and replaced both the federal Family Allowances and the Child Tax Credit. Several provincial and territorial programs have since been introduced, in addition to Quebec family allowances which already existed before 1993. To be eligible, a person must have the primary responsibility for the care and upbringing of one or more children under the age of 18. Most benefits are calculated by setting a maximum amount per family or per child and reducing that total by a certain percentage of the family's net income. The programs which were explicitly accounted for in the data for 1999 were: the federal basic benefit and National Child Benefit Supplement (together called the Canada Child Tax Benefit), the Newfoundland and Labrador Child Benefit, the Nova Scotia Child Benefit, the New Brunswick Child Tax Benefit, the New Brunswick Earned Income Supplement, the Quebec Allocation familiale, the Quebec Allocation à la naissance, the Ontario Child Care Supplement for Working Families, the Saskatchewan Child Benefit, the Alberta Family Employment Tax Credit, the BC Family Bonus, and the BC Earned Income Benefit.

Old Age Security (OAS) benefits

The Old Age Security (OAS) pension is targeted to Canadian residents aged 65 and over. OAS recipients who have little or no other income may also receive the federal Guaranteed Income Supplement (GIS); and their spouses, if aged 60 to 64 (and not yet eligible for OAS and GIS themselves), receive the Spouse's Allowance.

Canada Pension Plan (CPP) and Quebec Pension Plan (QPP) benefits

The CPP and QPP are compulsory contributory social insurance programs that provide a source of retirement income and protect workers and their families against loss of income due to disability or death.

Employment Insurance benefits

Employment Insurance is a federal program which includes the following types of benefits: regular unemployment benefits, sickness benefits, maternity and parental benefits, and benefits for persons taking approved training courses or participating in job creation or job-sharing projects. To qualify, the claimant must have ceased receiving employment income and have worked a minimum number of weeks or hours of insurable employment over the preceding period.

Social assistance

Social assistance covers many provincial and municipal income supplements to individuals and families. It is usually provided only after all other possible sources of support have been exhausted.

Workers' compensation benefits

Workers' compensation is provided to protect all full-time and part-time employees from loss of salary due to work accidents or occupational diseases and help them to pay their medical expenses and other costs.

Goods and Services Tax/Harmonized Sales Tax Credit

This credit was introduced in conjunction with the Goods and Services Tax in 1990. It is intended to offset the GST/HST for lower-income families and individuals. In Nova Scotia, New Brunswick and Newfoundland and Labrador, its name was changed to the Harmonized Sales Tax Credit in April 1997 when the administration of the tax was combined with provincial sales tax.

Provincial/territorial tax credits

Included here are refundable tax credits other than those for children (included with child tax benefits). Some are designed to help low-income individuals and families to pay property taxes, education taxes, rent and living expenses, and so on. Provincial sales tax credits such as the Quebec Sales Tax Credit and the Newfoundland and Labrador HST Credit are included. The Quebec abatement, although refundable, is not included here but rather with income taxes.

Other government transfers

This includes government transfers not included elsewhere, mainly any other non-taxable transfers. In SLID, these amounts are included with "Other income". This is partly because the coverage of any transfers not taxed through the income tax system is low. In survey interviews, there may be under-reporting of these transfers, which are mainly collected using an open question. Nonetheless, the types of transfers which have come under this heading include: training program

payments not reported elsewhere, the Veteran's pension, pensions to the blind and the disabled, regular payments from provincial automobile insurance plans (excluding lump-sum payments), and benefits for fishing industry employees.

Other income

This subtotal includes all items of market income not included elsewhere. Among them are support payments received (also called alimony and child support). The coverage of other items depends at least to some extent on the method of income data collection, whether from administrative income tax records or by interview. Those items that are included on line 130 of the T1 tax return are well covered. These include, but are not restricted to, retiring allowances (severance pay/termination benefits), scholarships, lump-sum payments from pensions and deferred profit-sharing plans received when leaving a plan, the taxable amount of death benefits other than those from CPP or QPP, and supplementary unemployment benefits not included in wages and salaries.

Income tax

Income tax is the sum of federal and provincial income taxes payable (accrued) for the taxation year. Income taxes include taxes on income, capital gains and RRSP withdrawals, after taking into account exemptions, deductions, non-refundable tax credits, and the refundable Quebec abatement. In the Survey of Labour and Income Dynamics, the data are either taken directly from administrative records or estimated based on aggregate data from administrative records, as this yields better results than the amounts reported by interview.

After-tax income

After-tax income is total income, which includes government transfers as defined here, less income tax. It may also be called income after tax.

Family definitions

Dwelling

In general terms, a dwelling is defined as a set of living quarters. A private dwelling is a separate set of living quarters with a private access. A collective dwelling may be institutional, communal or commercial in nature. Of the different types of collective dwellings, only communal dwellings are covered in the SLID.

Household

A household is defined as a person or group of persons residing in a dwelling.

Economic family

An economic family is defined as a group of two or more persons who live in the same dwelling and are related to each other by blood, marriage, common-law or adoption.

Unattached individual

An unattached individual is a person living either alone or with others to whom he or she is unrelated, such as roommates or a lodger.

Census family

The term "census family" corresponds to what is commonly referred to as a "nuclear family" or "immediate family". In general, it consists of a married couple or common-law couple with or without children, or a lone-parent with a child or children; furthermore, each child does not have his or her own spouse or child living in the household.

Persons "not in census families" are those living alone, living with unrelated individuals, or living with relatives but not in a husband-wife or parent-unmarried child (including guardianship child) relationship.

By definition, all persons who are members of a census family are also members of the same economic family.

Adults

Adults are defined in SLID as 16 or older as of December 31 of the reference year.

Family income

Family income is the sum of income of each adult in the family as defined above. Household income is likewise the sum of incomes of all adults in the household. Family and household membership is defined at a particular point in time, while income is based on the entire calendar year. The family members or "composition" may have changed during the reference year, but no adjustment is made to family income to reflect this.

SLID defines households and families according to the living arrangements on December 31 of the reference year.

Major income earner

This characteristic is important for the derivation of detailed family types. For each household and family, the major income earner is the person with the highest income before tax, with one exception; a child living in the same census family as his/her parent(s) cannot be identified as the major income earner of the census family (this does not apply to economic families).

For persons with negative total income before tax, the absolute value of their income is used, to reflect the fact that negative incomes generally arise from losses "earned" in the market place and are not meant to be sustained. In the rare situations where two persons have exactly the same income, the older person is the major income earner.

Table B
Classification of family types

Economic families (or Census families), two persons or more
- Elderly families
  - Married couples
  - All other elderly families
- Non-elderly families
  - Married couples without children
    - No earner
    - One earner
    - Two earners
  - Two-parent families with children
    - No earner
    - One earner
    - Two earners
    - Three or more earners
  - Married couples with other relatives
  - Lone-parent families
    - Male lone-parent families
    - Female lone-parent families
      - No earner
      - One earner
      - Two or more earners
  - Other non-elderly families

Unattached individuals (or persons not in Census families)
- Elderly male
  - Non-earner
  - Earner
- Elderly female
  - Non-earner
  - Earner

Non-elderly male
    Non-earner
    Earner
Non-elderly female
    Non-earner
    Earner

Within this classification, the following definitions apply:

Elderly family: The major income earner in the economic family is aged 65 or over.

Married couples/Spouses: Married couples include legally married, common-law and same-sex relationships where one of the spouses is the major income earner.

Children: A child or children (by birth, adopted, step, or foster) of the major income earner under age 18. Other relatives may also be in the family.

Lone-parent family: Includes at least one child as defined above. Families where the parent is 65 years or older are excluded.

Relative: A person related to the major income earner by blood, marriage, adoption or common-law.

Other relative: A person in the economic family who is not the major income earner nor his/her spouse or child under age 18.

Analytical concepts

Current dollars versus constant dollars

"Current dollars" are what we usually mean when we refer to a currency in the current time period. The term "constant dollars" refers to dollars of several years expressed in terms of their value ("purchasing power") in a single year, called the base year. This type of adjustment is done to eliminate the impact of widespread price changes. Current dollars are converted to constant dollars using an index of price movements. The most widely used index for household or family incomes, provided that no specific uses of the income are identified, is the Consumer Price Index (CPI), which reflects average spending patterns by consumers in Canada.

The following table shows the annual rates of the Consumer Price Index. To convert current dollars of any year to constant dollars, divide them by the index of that year and multiply them by the index of the base year you have chosen (remember that the numerator contains the index value of the year you want to move to). For example, using this index, $10,000 in 1997 would be $10,548 in 2000 constant dollars ($10,000 × 113.5/107.6 = $10,548).

Table C
Consumer Price Index, annual rates, 1992=100

1980	52.4	1990	93.3	2000	113.5
1981	58.9	1991	98.5	2001	116.4
1982	65.3	1992	100.0		
1983	69.1	1993	101.8		
1984	72.1	1994	102.0		
1985	75.0	1995	104.2		
1986	78.1	1996	105.9		
1987	81.5	1997	107.6		
1988	84.8	1998	108.6		
1989	89.0	1999	110.5		

Earner/Income recipient

An earner is a person who received income from employment (wages and salaries) and/or self-employment during the reference year. The term income recipient is generally used for someone who received a positive (or negative) amount of income of any given type.

Mean income (average income)

The mean or average income is computed as the total or "aggregate" income divided by the number of units in the population. It offers a convenient way of tracking aggregate income while adjusting for changes in the size of the population.

There are two drawbacks to using average income for analysis. First, since everyone's income is counted, the mean is sensitive to extreme values: unusually high income values will have a large impact on the estimate of mean income, while unusually low ones, i.e. highly negative values, will drive it down. (See also "Recipients versus non-recipients" and "Negative values".) Secondly, it does not give any insight into the allocation of income across members of the population. For this, measures such as percentiles or Gini coefficients may be used.

Recipients versus non-recipients (zero values)

For every table showing average incomes, it must be kept in mind whether non-recipients of that type of income are included or excluded from the population. In the case of total family income, the difference of including or excluding units with zero income is small since there are very few such families. However, if one is interested in the average amount of individual self-employment earnings, the value will be quite different if one includes those persons who were not self-employed. Zero values are included in all tables focussing on the three main

income concepts (market, total, or after-tax income), government transfers or taxes.

Negative values

Negative income amounts can arise in two ways: net losses from self-employment (expenses exceed receipts), or net investment losses (losses exceed gains). As with zero values, negative values can have a large impact on results. In general, the published income tables treat negative values no differently than positive values, but there are a few exceptions: for the calculation of both Gini coefficients and the low income gap, negative values are converted to zeroes; and in the derivation of the major income earner of a family or household, the absolute value is used instead (see "Major income earner" under "Family definitions").

Percentiles

Income percentiles like quintiles and deciles are a convenient way of categorizing units of a given population from lowest income to highest income for the purposes of drawing conclusions about the relative situation of people at either end or in the middle of the scale. Rather than using fixed income ranges, as in a typical distribution of income, it is the fraction of each population group that is fixed.

First, all the units of the population, whether individuals or families, are ranked from lowest to highest by the value of their income of a specified type, such as after-tax income. Then, the ranked population is divided into five groups of equal numbers of units, called quintiles. Analogously, dividing the population ranked by income into ten groups, each comprising the same number of units, produces deciles.

Most analyses should be carried out on the people of different percentiles within a distribution. Care should be taken in making comparisons between quintiles that resulted from different distributions, because any difference in either the population or the income concept used to rank units could have a large effect. It is probable that both the income ranges represented by each quintile and the people making up each quintile will be different.

Median income

The median income is the value for which half of the units in the population have lower incomes and half have higher incomes. To derive the median value of income, units are ranked from lowest to highest according to their income and then separated into two equal-sized groups. The value that separates these groups is the median income. It corresponds to the 50th percentile.

Because the median corresponds exactly to the mid-point of the income distribution, it is not, contrary to the mean, affected by extreme income values. This is a useful feature of the median, as it allows one to abstract from unusually high values held by relatively few people.

Since income distributions are typically skewed to the left - that is, concentrated at the low end of the scale - median income is usually lower than mean income.

Implicit rate of government transfers or taxes

The implicit rate of either transfers or taxes, as the case may be, is a way of showing the relative importance of transfers received or taxes paid for different families or individuals. This concept is similar, but not identical, to the effective rate of taxation. For a given individual or family, the effective rate is the amount of transfers/taxes expressed as a percentage of their income, usually market income, total income, or after-tax income. The implicit rate for a given population is the average (or aggregate) amount of transfers/taxes expressed as a percentage of their average (or aggregate) income.

Family size adjustment (equivalence scale)

When comparing family incomes to study such things as income adequacy or socio-economic status, one often wants to take the family size into account. Basically stated, the income amount itself is not sufficient to understand a family's financial well-being without knowing how many people are sharing it. Two approaches have been used to help with the analysis of family income. One is to produce data by detailed family types, so that within a given family type, differences in family size are not significant. In fact, many income measures have been crossed by detailed family types in the published tables.

The other way to take family size into account is to adjust the income amount, for the purposes of analysis only. The major challenge of this approach is to select an appropriate adjustment factor. While there is no single best method, it is still better to apply some kind of adjustment factor rather than no adjustment at all.

The simplest method is to use per capita income, that is, to divide the family income by the family size. A limitation of per capita income, however, is that it tends to underestimate economic well-being for larger families as compared to smaller families. This is due to the fact that it assumes equal living costs for each member of the family, but some costs, primarily those related to shelter, decrease proportionately with family size (they may also be lower for children than for adults). For example, the shelter costs for an adult married couple with no children are arguably not much more than those for an adult living alone.

To take such economies of scale into account, it is common to use an "equivalence scale" to adjust family incomes. Instead of implicitly assuming

equal costs for additional family members as the per capita approach does, the equivalence scale is a set of decreasing factors assigned to the first member, the second member, and so on. Dividing the income value by the sum of the factors assigned to each member derives the adjusted income amount for the family.

There is no single equivalence scale in use in Canada. The one used in the published income tables and in concepts such as the Low Income Measure (LIM) has, however, achieved a high degree of acceptance. In this equivalence scale, the factors are as follows:

- the oldest person in the family receives a factor of 1.0;
- the second oldest person in the family receives a factor of 0.4;
- all other family members aged 16 and over each receive a factor of 0.4;
- all other family members under age 16 receive a factor of 0.3.

For example, this translates into a total factor for dividing income of just 1.4 for a married couple instead of 2.0 (the family size). Such a family with total income of $56,000 would be considered to have a standard of living equivalent to an adult living alone with a total income of $40,000, as compared to an adult with $28,000 when calculated on a per capita basis.

Gini coefficient

The Gini coefficient measures the degree of inequality in an income distribution. Gini coefficients are published for a variety of income measures such as market income, total income and after-tax income, and are used to compare the uniformity of income allocation between different income concepts across different populations or within the same population over time.

Values of the Gini coefficient can range from 0 to 1. A value of zero indicates income is equally divided among the population with all units receiving exactly the same amount of income. At the opposite extreme, a Gini coefficient of 1 denotes a perfectly unequal distribution where one unit possesses all of the income in the economy. A decrease in the value of the Gini coefficient can, by and large, be interpreted as reflecting a decrease in inequality, and vice versa. As a rough rule of thumb when using data from SLID at the Canada level, a difference of 0.01 or more between two Gini coefficients is considered statistically significant.

Low income definitions

Low income cutoff (LICO)

Low income cutoffs (LICOs) are established using data from the Family Expenditure Survey, now known as the Survey of Household Spending. They convey the income level at which a family may be in straitened circumstances because it has to spend a greater proportion of its income on necessities than the average family of similar size. Specifically, the threshold is defined as the income below which a family is likely to spend 20 percentage points more of its income on food, shelter and clothing than the average family. There are separate cutoffs for seven sizes of family – from unattached individuals to families of seven or more persons – and for five community sizes – from rural areas to urban areas with a population of more than 500,000.

Calculation of low income cutoffs

The first step in the production of a set of low income cutoffs is to calculate the average proportion of income that a family spends on food, shelter and clothing. The 1992 Family Expenditure Survey found that, on average, families spend 44% of their after-tax income (and 35% of their total "before-tax" income) on these necessities. Then, 20 percentage points are added, giving 64% of after-tax income. This is done on the grounds that a family spending more than this proportion of its income on necessities is significantly worse off than the average family. The final step is to look at the distribution of income by expenditure and determine, using a regression line, the level of income at which a family tends to spend 20 percentage points more than the average on the necessities of food, shelter and clothing.

Updating and rebasing the low income cutoffs

There are two reference years that play a part in the calculation of a set of low income cutoffs: the base year and the income reference year. The base year supplies the average spent on food, shelter and clothing. This percentage is used to derive a set of cutoffs that are suitable for use with income data from that year. Cutoffs for other income reference years may be obtained by applying the corresponding Consumer Price Index (CPI) inflation rate to the basic set of cutoffs.

Using the CPI to update the cutoffs takes inflation into account, but does not reflect any changes that might occur in the average spending on necessities. In the past, Statistics Canada has developed a new set of cutoffs after each Family Expenditure Survey. These are referred to as 'bases' because the average spending on necessities in that base year drives the calculation of the cutoffs. The two most recent base years are 1992 and 1986. Cutoffs based on 1992 are most

commonly applied by data users, and are available for the income reference years from 1980 onwards.

Low income rate

Low income rates can be calculated for persons or for families. In either case, the income that is compared to the cutoff is the income of the entire economic family. "Persons in low income" should be interpreted as persons who are part of low income families including persons living alone whose income is below the cutoff. Similarly, "children in low income" means "children who are living in low income families". In other words, all members of an economic family have the same low income status, but they are counted separately when person-based low income rates are calculated.

To calculate the low income rates, the family size and community size are used to find the appropriate cutoff. Then the family income is compared to that cutoff. If a family low income rate is being calculated, then the family is counted as being in low income if its income is less than the cutoff. If a person low income rate is being calculated, then all persons in the family are counted as being in low income if the family income is less than the cutoff.

Use of after-tax and before-tax LICOs

The average portion of income that families spend on food, shelter and clothing, which figures prominently in the low income cutoffs, is undoubtedly a useful gauge of economic well-being no matter which income concept is used. The choice of after-tax income or total income – or even market income for that matter – depends on whether one wants to take into account the added spending power that a family gets from receiving government transfers and its reduced spending power from paying taxes.

In the past, Statistics Canada has produced two sets of low income cutoffs and corresponding rates – those based on total income (i.e. income including government transfers, before the deduction of income taxes) and those based on after-tax income.

The choice to highlight after-tax rates was made for two main reasons. First, income taxes and transfers are essentially two methods of income redistribution. The before-tax rates only partly reflect the entire redistributive impact of Canada's tax/transfer system, by including the effect of transfers but not the effect of income taxes. Second, since the purchase of necessities is made with after-tax dollars, it is logical to use people's after-tax income to draw conclusions about their overall economic well-being.

A note about the calculation of before-tax versus after-tax low income cutoffs: the derivation of each set of cutoffs is done independently. There is no simple

relationship, such as the average amount of taxes payable, that distinguishes the two levels. Instead, the entire calculation of cutoffs is done twice – both on a before-tax basis and on an after-tax basis.

Differences in after-tax rates and before-tax rates

After-tax low income cutoffs, and the resulting after-tax rates, have been published back to 1980. The number of people falling below the cutoffs has been consistently lower on an after-tax basis than on a before-tax basis. This result may appear inconsistent at first glance, since incomes after tax cannot be any higher than they are before tax, considering that all transfers, including refundable tax credits, are included in the definition of "before-tax" total income. However, with a relative measure of low income such as the LICO, this result is to be expected with any income tax system which, by and large, taxes those with more income at a higher rate than those with less. "Progressive" tax rates, as they are often called, make the distribution of income more compressed. Therefore, some families that are in low income before taking taxes into account are relatively better off and are not in low income on an after-tax basis.

Low income gap

The low income gap, previously called "low income deficiency", is the amount that a low income family falls short of the relevant low income cutoff. For the calculation of this gap, negative incomes are treated as zero.

For example, a family with an income of $15,000 and a relevant low income cutoff of $20,000 would have a low income gap of $5,000. In percentage terms this gap would be 25%. The average gap for a given population, whether expressed in dollar or percentage terms, is the average of this value as calculation for each unit.

Market basket measure (MBM)

Human Resources Development Canada has been collaborating with the provincial and territorial ministries of social services to develop a "Market Basket Measure" (MBM). The approach is to cost out a basket of necessary goods and services including food, shelter, clothing and transportation, and a multiplier to cover other essentials. The results would define levels of income needed to cover the cost of the basket.

The same argument that can be made for using after-tax low income rates can be made for using after-tax income to compare to the MBM thresholds. That is, a measure of well-being should take into account what is actually available to spend. The income concept that has been proposed for comparisons with the MBM thresholds goes even further than after-tax income by also removing other non-discretionary expenses such as support payments, work-related child care

costs and employee contributions to pension plans and to Employment Insurance. Statistics Canada has started to collect some of the data necessary to produce rates based on the Market Basket Measure.

This type of measure is often called an "absolute" measure, even though there is clearly judgement involved in specifying the contents of the basket of goods and services. Nevertheless, the line is absolute in the sense that it does not depend directly on the distribution of income.

On Poverty and Low Income

Ivan P. Fellegi
Chief Statistician of Canada

Recently the news media have provided increasing coverage of Statistics Canada's low income cutoffs and their relationship to the measurement of poverty. At the heart of the debate is the use of the low income cutoffs as poverty lines, even though Statistics Canada has clearly stated, since their publication began over 25 years ago, that they are not. The high profile recently given to this issue has presented Statistics Canada with a welcome opportunity to restate its position on these issues.

Many individuals and organizations both in Canada and abroad understandably want to know how many people and families live in "poverty", and how these levels change. Reflecting this need, different groups have at different times developed various measures which purported to divide the population into those who were poor and those who were not.

In spite of these efforts, there is still no internationally-accepted definition of poverty - unlike measures such as employment, unemployment, gross domestic product, consumer prices, international trade and so on. This is not surprising, perhaps, given the absence of an international consensus on what poverty is and how it should be measured. Such consensus preceded the development of all other international standards.

The lack of an internationally-accepted definition has also reflected indecision as to whether an international standard definition should allow comparisons of well-being across countries compared to some international norm, or whether poverty lines should be established according to the norms within each country.

The proposed poverty lines have included, among others, relative measures (you are poor if your means are small compared to others in your population) and absolute measures (you are poor if you lack the means to buy a specified basket of goods and services designated as essential). Both approaches involve judgmental and, hence, ultimately arbitrary choices.

In the case of the relative approach, the fundamental decision is what fraction of the overall average or median income constitutes poverty. Is it one-half, one-third, or some other proportion? In the case of the absolute approach, the number of individual judgements required to arrive at a poverty line is far larger. Before anyone can calculate the minimum income needed to purchase the "necessities" of life, they must decide what constitutes a "necessity" in food, clothing, shelter and a multitude of other purchases, from transportation to reading material.

The underlying difficulty is due to the fact that poverty is intrinsically a question of social consensus, at a given point in time and in the context of a given country. Someone acceptably well off in terms of the standards in a developing country might well be considered desperately poor in Canada. And even within the same country, the outlook changes over time. A standard of living considered as acceptable in the previous century might well be viewed with abhorrence today.

It is through the political process that democratic societies achieve social consensus in domains that are intrinsically judgmental. The exercise of such value judgements is certainly not the proper role of Canada's national statistical agency which prides itself on its objectivity, and whose credibility depends on the exercise of that objectivity.

In Canada, the Federal/Provincial/Territorial Working Group on Social Development Research and Information was established to create a method of defining and measuring poverty. This group, created by Human Resources Development Canada and social services ministers in the various jurisdictions, has proposed a preliminary market basket measure of poverty - a basket of market-priced goods and services. The poverty line would be based on the income needed to purchase the items in the basket.

Once governments establish a definition, Statistics Canada will endeavour to estimate the number of people who are poor according to that definition. Certainly that is a task in line with its mandate and its objective approach. In the meantime, Statistics Canada does not and cannot measure the level of "poverty" in Canada.

For many years, Statistics Canada has published a set of measures called the low income cutoffs. We regularly and consistently emphasize that these are quite different from measures of poverty. They reflect a well-defined methodology that identifies those who are substantially worse off than the average. Of course, being significantly worse off than the average does not necessarily mean that one is poor.

Nevertheless, in the absence of an accepted definition of poverty, these statistics have been used by many analysts to study the characteristics of the relatively worst off families in Canada. These measures have enabled us to report important trends, such as the changing composition of this group over time. For example, 20 to 30 years ago the elderly were by far the largest group within the "low income"

category, while more recently lone-parent families headed by women have grown in significance.

Many people both inside and outside government have found these and other insights to be useful. As a result, when Statistics Canada carried out a wide-ranging public consultation a decade ago, we were almost unanimously urged to continue to publish our low income analyses. Furthermore, in the absence of a generally accepted alternative methodology, the majority of those consulted urged us to continue to use our present definitions.

In the absence of politically-sanctioned social consensus on who should be regarded as "poor", some people and groups have been using the Statistics Canada low income lines as a de facto definition of poverty. As long as that represents their own considered opinion of how poverty should be defined in Canada, we have no quarrel with them: all of us are free to have our own views. But they certainly do not represent Statistics Canada's views about how poverty should be defined.

5. Guidelines for applying weights

The microdata on the public use file are unweighted. It is the responsibility of data users to apply the appropriate weights in any estimates they wish to produce. If proper weights are not used, the results derived from the microdata cannot be considered to be representative of the survey population, and will not correspond to those that would be produced by Statistics Canada. The weights are provided as variables under "Sample control". On the SLID PUMF, the weight variable is named ICSWT26.

6. Guidelines for release (data quality and rounding)

Microdata users should apply the rules for assessing data quality, below, to all estimates they produce, and retain only those that satisfy the release criteria. Estimates that do not satisfy the release criteria are not reliable.

Introduction

The guidelines for release and publication make use of the concept of sampling variability to determine whether the estimates obtained from the microdata are reliable. Sampling variability is the error in the estimates caused by the fact that we survey a sample rather than the entire population. The concept of standard error and the related concept of coefficient of variation and confidence interval provide an indication of the magnitude of the sampling variability.

The standard error and coefficient of variation do not measure any systematic biases in the survey data which might affect the estimate. Rather, they are based on the assumption that the sampling errors follow a normal probability distribution.

Subject to this assumption, it is possible to estimate the extent to which different samples that have the same design and the same number of observations would give different results. This indicates the margin of error that is likely to be included in the estimates derived from our single sample.

For a more complete description of the measures of sampling variability, see A. Satin and W. Shastry, Survey Sampling: A Non-Mathematical Guide, Statistics Canada, Catalogue 12-602E.

Minimum sizes of estimates for release

In general, the smaller the sample, the greater the sampling variability. Consequently, estimates of small population subgroups are less reliable than estimates of large population subgroups. The minimum allowable sizes of estimates, also called the release cut-offs, are a quick rule for determining whether an estimate can be released, before applying the more rigorous test that uses the coefficient of variation. The release cut-offs are calculated specifically for the Survey of Labour and Income Dynamics based on the sample size and the sample design.

Both the cut-offs for the unweighted count and the weighted count must be satisfied:

- Unweighted count. The number of observations must be at least 25. If the unweighted count is less than 25, then the weighted estimate should not be released regardless of the value of its coefficient of variation.

- Weighted count. The following table shows the minimum levels for weighted counts by geographical area. For an estimate obtained from a numeric variable, such as an average, the same cut-off is compared with the estimate of the population to which the value applies (for example, the total number of hours worked by paid workers applies to the population of paid workers).

Table D
Release cut-offs based on the weighted estimate/count

CANADA	13,000
Newfoundland and Labrador	2,500
Prince Edward Island	1,500
Nova Scotia	4,000
New Brunswick	2,500
Quebec	14,000
Ontario	14,500
Manitoba	6,500
Saskatchewan	2,500
Alberta	6,000
British Columbia	11,000

Estimates of provincial aggregates and means

When producing estimates for provincial aggregates and means it should be noted that for a small number of records, province of residence has been suppressed. This will result in a small bias in provincial estimates.

Rounding guidelines

Once it has been determined whether the results obtained are reliable, the level of rounding indicates the level of precision that the data can actually support. The following guidelines for rounding should be used:

- All estimates should be rounded so there are no more than three significant digits.
- Estimates of population sub-groups should be rounded to at least the nearest hundred units.
- Rates and percentages should be rounded to at least one decimal point.

Note that all calculations are to be derived from their unrounded components, and then rounded using the normal rounding technique.

In normal rounding, if the first or only digit to be dropped is 0 to 4, the last digit to be retained is not changed. If the first or only digit to be dropped is 5 to 9, the last digit to be retained is raised by one. For example, in normal rounding to the nearest 100, the estimate 49,448 would be rounded down to 49,400 and an estimate of 49,252 would be rounded up to 49,300. The figure 1.78% would be rounded to 1.8%.

Hypothesis tests provided by statistical software packages

Microdata users should be aware that the results of hypothesis tests (such as the p values accompanying t statistics or Pearson statistics) that are provided automatically by most standard statistical software packages are incorrect for data provided by surveys with a complex survey design, such as SLID. Such packages calculate these test results under the assumption of simple random sampling. That is, they do not take into account the special sample design features of SLID such as stratification, clustering, and unequal selection probabilities. While many of the standard packages can account for the unequal selection probabilities in the production of estimates by allowing the use of weights, these packages do not properly take the sample design into account when producing variance estimates that form part of most test statistics.

To perform hypothesis tests, a two-step method can be employed with the standard statistical software to form the test statistics. First, estimate the characteristics of interest (total or mean) using the weights provided on the microdata file. Second, obtain approximate variance estimates of these characteristics by rerunning the same software procedure as that used for producing the characteristic estimates but using a scaled weight that consists of the original weight divided by the average of the original weights of all the observations being used in your computations. The standard error can be derived by using the estimate and the rough estimate of the variance. These quantities (estimate, variance, standard error) can then be combined to form test statistics. It must be noted that this method provides only rough approximations to the variance.

It should be noted that users of the SLID PUMF cannot readily obtain better design-based variance estimates through the use of statistical software specifically designed for survey data. This is because the design information required by these software packages is not currently available on the SLID data file due to confidentiality considerations. However, better variance estimates can be produced by Statistics Canada on a cost-recovery basis.

7. Confidentiality of the public-use microdata

The production of a public-use microdata file includes many safeguards to prevent the identification of any one person. Longitudinal surveys are faced with an extra challenge when it comes to ensuring confidentiality, because data are collected for the same people for several years. For this reason, Statistics Canada plans to release only cross-sectional files from SLID. The number of topics covered in SLID also contributes to the amount of processing required specifically to ensure confidentiality. Confidentiality of the public-use file is ensured mainly by reducing information, i.e. deleting whole variables or suppressing or collapsing some of their detail.

SLID uses a number of techniques to ensure confidentiality:

- The SLID public-use file is comprised of a sample of the households randomly selected from the full SLID sample.
- All the variables which would permit direct identification of individuals are, of course, deleted from the file. This includes name, telephone number, and other data used for collection purposes;
- Collapsing categories. This is applied to categorical (i.e. qualitative) variables.
- Top and bottom coding. Very high and very low values usually are rare or unique in the population. Such extreme values are replaced with the value of an upper or lower limit.
- Rounding. Some variables, particularly monetary values, are rounded.
- Suppression and modification of characteristics was done while preserving integrity of the file for the purpose of producing precise and accurate statistics.
- Imputed records and variables on the file are not identified as such.
- Addition of "noise" (perturbation). Numeric values may have been raised or reduced by unequal amounts and proportions in a random-like fashion (addition of "noise"), while maintaining data integrity for the purpose of producing precise and accurate statistics.

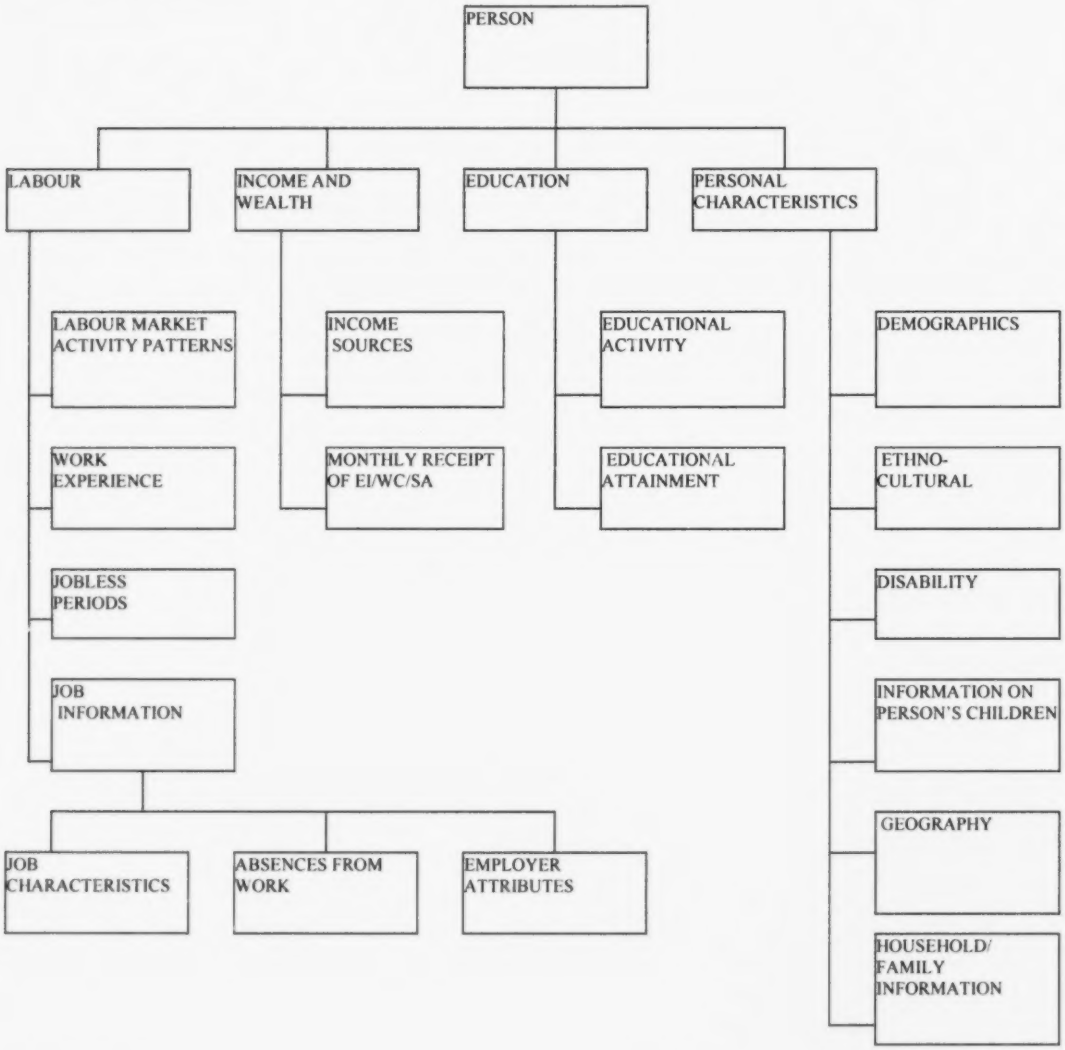
8. Sources, methods and estimation procedures

Survey content

SLID was designed to capture changes in the economic well-being of individuals and families over time and the determinants of labour market and income changes. The survey supports analysis on transitions into and out of the labour force associated with the life cycle or with the business cycle; on the impact of family events on labour market activity and remuneration; on the determinants of income instability; on what triggers shifts into and out of low income and on changes in the composition of income through time. Since SLID additionally carries a broad selection of human capital variables, it is also used for studies of such topics as gender wage and earnings gaps.

The major content themes of SLID are illustrated in the following chart.

Chart A
Organization of content



Survey universe

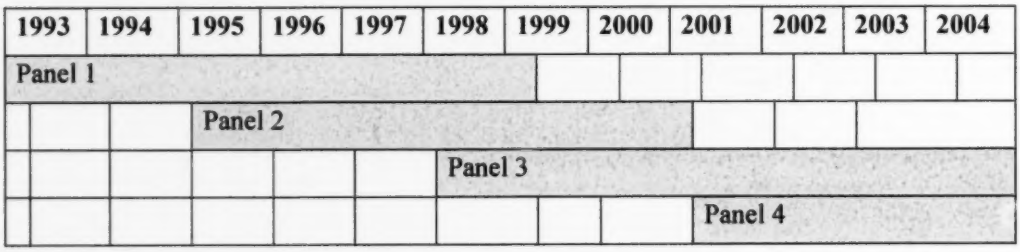
SLID covers all individuals in Canada, excluding residents of the Yukon, the Northwest Territories and Nunavut; residents of institutions and persons living on Indian reserves. Overall, these exclusions amount to less than 3 percent of the population.

The sample

The samples for SLID are selected from the monthly Labour Force Survey (LFS) and thus share the latter's sample design. The LFS sample is drawn from an area frame and is based on a stratified, multi-stage design that uses probability sampling. The sample is composed of six independent samples. These samples are called rotation groups because each month one sixth of the sample (or one rotation group) is replaced.

The SLID sample is composed of two panels. Each panel consists of two LFS rotation groups and includes roughly 15,000 households. A panel is surveyed for a period of six consecutive years. A new panel is introduced every three years. Thus two panels are always overlapping. The following diagram illustrates how and when panels overlap.

Chart B
Overlapping design of SLID sample



Data collection

For each sampled household in SLID, up to 12 interviews are conducted over a six-year period. Every year in January, interviewers collect information regarding respondents' labour market experiences during the previous calendar year. Information on educational activity and family relationships is also collected at that time. The demographic characteristics of family and household members represent a snapshot of the population as of the end of each calendar year.

Every May information on income is collected from the same sampled households. The income interview is deferred until May to take advantage of income tax time when respondents are more familiar with their income situation. The reference period for income is the previous calendar year.

To reduce response burden, respondents can give Statistics Canada permission to use their T1 tax form information for the purposes of SLID. Over 80 percent of SLID's respondents give their consent to the use of their tax records. They are not contacted in May for the income interview.

The SLID interviews are conducted over the telephone using computer assisted interviewing (CAI). The interviewer reads the questions as they appear on the computer screen and keys in the reported information. Skip patterns and edits are built into the collection software, allowing interviewers to immediately detect and resolve response inconsistencies. Collection of date-related information (e.g., employment spells, jobless spells, interruption of work) is greatly improved by the use of such an interactive data capture technique. Another advantage of the CAI technology is the feeding back of details from the previous interview assisting the respondents to recall past events.

Proxy response is accepted in SLID. This procedure allows one household member to answer questions on behalf of any or all other members of the household, provided he or she is willing to do so and is knowledgeable.

Data quality

There are two types of errors inherent to sample survey data, namely, sampling errors and non-sampling errors. The reliability of survey estimates depends on the combined impact of sampling and non-sampling errors.

Sampling errors

Sampling errors occur because inferences about the entire population are based on information obtained from only a sample of the population. The results are usually different from those that would be obtained if information were collected from the whole population. Errors due to the extension of conclusions based on the sample to the entire population are known as sampling errors. The sample design, the variability of the population characteristics measured by the survey, and the sample size determine the magnitude of the sampling error. In addition, for a given sample design, different methods of estimation will result in sampling errors of different sizes.

Standard error and coefficient of variation

A common measure of sampling error is the standard error (SE). The standard error measures the degree of variation introduced in estimates by selecting one particular sample rather than another of the same size and design. The standard error may also be used to calculate confidence intervals associated with an estimate (Y). Confidence intervals are used to express the precision of the estimate. It has been demonstrated mathematically that, if the sampling were repeated many times, the true population value would lie within the $Y \pm 2SE$ confidence interval 95 times out of 100 and within the narrower confidence interval defined by $Y \pm SE$, 68 times out of 100. Another important measure of sampling error is given by the coefficient of variation, which is computed as the estimated standard error as a percentage of the estimate Y (i.e. 100 x SE /Y).

To illustrate the relationship between the standard error, the confidence intervals and the coefficient of variation, let us take the following example. Suppose that the estimated average income from a given source is $10,000, and that its corresponding standard error is $200. The coefficient of variation is therefore equal to 2%. The 95% confidence interval estimated from this sample ranges from $9,600 to $10,400, i.e. $10,000 ± $400. This means that with a 95% degree of confidence, it can be asserted that the average income of the target population is between $9,600 and $10,400.

The bootstrap approach is used for the calculation of the standard errors of the SLID estimates. For more information on standard errors and coefficients of variation, refer to the Statistics Canada publication Methodology of the Canadian Labour Force Survey (Catalogue 71-526-XPB).

Standard errors and coefficients of variation of the estimates produced from this file are available on request. An approximate method is provided in Section 6 of this document.

Non-sampling errors

Non-sampling errors generally result from human errors such as inattention, misunderstanding or misinterpretation. The impact of randomly occurring errors over a large number of observations will be minimal. Errors occurring systematically can, on the other hand, have a major impact on the reliability of estimates. Considerable time and effort is invested into reducing non-sampling errors in SLID.

Non-sampling errors may arise from a variety of sources such as coverage, response, non-response and processing errors.

Coverage error arises when sampling frame units do not exactly represent the target population. Units may have been omitted from the sampling frame (undercoverage), or units not in the target population may have been included (overcoverage), or units may have been included more than once (duplicates). Undercoverage represents the most common coverage problem.

Slippage is a measure of survey coverage error. It is defined as the percentage difference between control totals (Census population projections) and weighted sample counts. Slippage rates for household surveys are generally positive because some people who should be enumerated are missed. According to the numbers below, in 1999, SLID covered 87.98% of its target population.

Table E
Slippage rates in SLID

Year	1999	2000
Canada (%)	12.02	12.64

Rates are also available upon request for sex, province and age groupings.

Response errors may be due to many factors, such as faulty questionnaire design, interviewers' or respondents' misinterpretation of questions, or respondents' faulty reporting. Great effort is invested in SLID to reduce the occurrence of response error. Measures undertaken to minimize response errors include the use of highly-skilled and well-trained interviewers, and supervision of interviewers to detect misinterpretation of instructions or problems with the questionnaire design. Response error can also be brought about by respondents who, willingly or not, provide inaccurate responses.

Income data are especially prone to misreporting, as income is a sensitive issue and includes many items with which respondents are not always familiar. To obtain more accurate information, income data for SLID are collected after the income tax "season" when respondents are more familiar with their tax records. Respondents receive information about the income interview prior to the interviewer's telephone call. This gives them time to consult documents and have information available at the time of the interview. For respondents who grant Statistics Canada permission to access their tax files (the majority of respondents), SLID collects income data directly from administrative files. This procedure reduces misreporting of income in the SLID.

Non-response errors occur to some extent in any survey for reasons such as household members being on vacation during the interview period or refusing to supply requested information, despite attempts to obtain complete response from sampled units. For these individuals, the missing data are imputed either explicitly by assigning data to each non-respondent on the basis of a similar respondent record, or implicitly by redistributing the weight of the non-respondent individual to other responding individuals. The bias introduced by non-response increases with the differences between respondent and non-respondent characteristics. Methods employed to compensate for non-response make use of information available for both respondents and non-respondents in an attempt to minimize this bias.

Processing errors can occur at various stages in the survey: data capture, editing, coding, weighting or tabulation. The computer-assisted collection method used for SLID reduces the chance of introducing capture errors because checks for consistency and completeness of the data are built into the computer application. To minimize coding, weighting or tabulation errors, diagnostic tests are carried

out periodically. These tests include comparisons of results with other data sources.

Cross-sectional representativeness of SLID

Each longitudinal sample, or "panel" in SLID initially constitutes a representative cross-sectional sample of the population. However, because the real population changes each year, whereas by design the longitudinal sample does not, the sample must be modified to properly reflect these changes to the composition of the population. This is done by adding to the sample all new people in the population who are found to be living with the initial respondents (and likewise dropping them from the sample if they leave at later time-points). Conversely, any original respondents who leave the target population (by moving abroad, into institutions, etc.) are given a zero weight for cross-sectional purposes. In this way, the cross-sectional sample, composed of the original respondents minus those who left the target population plus those who have entered it, is virtually fully representative of the population at each subsequent time-point. The missing group is composed of persons who have newly entered the target population and are not living with anyone who was in the target population when the most recent panel was selected. Since SLID introduces a new panel every three years, however, this group is quite small.

Response rates

High response rates are essential for the data quality of any survey and thus considerable effort is invested to encourage effective participation from SLID respondents.

The response rates are relatively high in SLID. SLID's cross-sectional rate of response varied from a low of 79.2% in reference year 2000 to a high of 86.0% in reference year 1996. The response rate is based on household response in SLID. For purposes of calculating cross-sectional response rates in SLID, households are defined according to the January household composition. The calculation of the response rate at the household level is based on the response codes for individuals in the household, including both longitudinal respondents and cohabitants. A respondent household is defined as a household that has at least one respondent individual. An individual is defined as a respondent if he or she responded to either the labour or the income interview.

Respondent households are divided into completely respondent households and partially respondent households. Partially respondent households are weighted and the missing income data in these households are imputed.

Table F
Response rate in and SLID (1996-2000)

Year	1996	1997	1998	1999	2000
Response Rate (%)	86.0	84.1	82.8	82.7	79.2

Imputation for non-response

In some cases, income data are imputed in SLID using a "nearest neighbour" approach. This method involves identifying another individual with certain similar characteristics, who becomes the "donor" for the imputed value.

SLID also uses other imputation techniques. In fact, the primary method employed for imputing income data in this survey is to use the previous year's data, updated for any changes in circumstances. Only in the absence of such data are income figures imputed using the "nearest neighbour" technique in SLID.

Amounts received through government programs such as the Child Tax Benefits, the Goods and Services/Harmonized Sales Tax Credit, the Guaranteed Income Supplement, are derived from other information collected by the survey. Data obtained from the tax route are considered complete and thus require no imputation.

9. SLID content

The SLID content organization is presented earlier in this document. Themes are organized under the topics of labour, income and wealth, education, and personal characteristics, including selections of the variables they contain. This section provides more detail on the content of SLID by content theme. Variables appearing on the public use file are marked with an asterisk *.

I. Labour

Nature and pattern of labour market activities

- major activity during year *
- spells of employment and unemployment (start and end dates, durations)
- monthly labour force status *
- total weeks of employment, unemployment and inactivity by year *
- multiple job-holding spells
- work absence spells

Work experience

- years of full-time and part-time employment
- years of experience in full-time, full-year equivalents *

Characteristics of jobless spells

- job search during spell
- dates of search spells
- desire for employment
- reason for not looking

Job characteristics (all characteristics updated each year and dates of changes recorded; collected for up to six jobs per year)

- start and end dates, first date ever worked for this employer
- wages *
- work schedule (hours and type) *
- benefits *
- union membership *
- occupation *
- supervisory and managerial responsibilities
- class of worker *
- tenure
- how job was obtained
- reason for job separation

Characteristics of work absences lasting one or more weeks (collected on first and last absence each year, for each employer)

- absence dates
- reason
- paid or unpaid

Employer attributes

- industry *
- firm size *
- public or private sector *

II. Income and wealth

Personal income

- annual information on 15 income sources *
- total income *
- taxes paid *
- after tax income *

Receipt of compensation (whether benefits were received from each source and, if so, in which months)

- Employment Insurance * - yes/no only on PUMF
- Social Assistance * - yes/no only on PUMF
- Workers' Compensation * - yes/no only on PUMF

III. Education

Educational activity

- enrolled in a credit program, months attended
- type of institution *
- full-time or part-time student *
- certificates received (if applicable) *

Educational attainment (updated annually)

- years of schooling *
- degrees and diplomas *
- major field of study

IV. Personal characteristics

Demographics

- year of birth / age *
- sex *
- duration of current marital status
- year/age at first marriage

Ethno-cultural

- ethnic background
- member of an Employment Equity designated group
- mother tongue
- date of immigration*
- country of birth
- parents' schooling and place of birth

Activity limitation

- annual information on activity limitations and their impact on working
- satisfaction with work

Information on person's children

- number of children born, raised *
- year and person's age when first child born

Geography and geographic mobility

- economic region or census metropolitan area of current residence
- size of community *
- moved during year
- move dates
- reason for move
- nature of move (full household/household split)

Household and economic family and census family information (annual summary information, e.g., size, type)

- key characteristics of other individuals in household/family (e.g., age, sex, relationship, income, annual hours worked)
- relevant low-income cutoff
- family events (marriage, separation, death, birth)
- dwelling type and tenure *

10. Related products and services

Canadian Statistics on the Internet

The following data are available, free of charge, on Statistics Canada's website (www.statcan.ca):

- Average Market Income by Selected Family Types, Canada
- Average Total Income by Selected Family Types, Canada
- Average After-Tax Income by Selected Family Types, Canada
- Government Transfers and Income Tax, Canada
- Persons in Low Income Before Tax, Canada
- Persons in Low Income After Tax, Canada

The menu path to download the above-listed tables is "Canadian Statistics", then "The People", followed by "Families, Households and Housing" and "Income".

- Average Earnings by Sex and Work Pattern, Canada
- Estimated numbers of Earners by Sex and Work Pattern, Canada

The menu path to download the above-listed tables is "Canadian Statistics", then "The People", followed by "Labour, Employment and Unemployment" and "Earnings".

Income in Canada (electronic version) 75-202-XIE ($34)

An electronic version of the present publication is available on Statistics Canada's website (www.statcan.ca).

The menu path to download the electronic version is "Our Products and Services", then "Browse our Internet Publications (for sale)", followed by "75-202-XIE, Income in Canada".

Income Trends in Canada, 13F0022XCB ($195)

This annual CD-ROM, which includes over 2 million data points, is the complement to Income in Canada. It provides historical trends starting in 1980, for Canada, the provinces and 15 metropolitan areas. The data are presented in Beyond 20/20 TM format that allows users to easily view trends, create tables and chart income.

Longitudinal data from the Survey of Labour and Income Dynamics (SLID)

SLID is a longitudinal survey – the same people are interviewed from one year to the next for a period of six years – that began collecting data with the 1993 reference year.

SLID includes a large selection of variables that capture transitions in Canadians' jobs, income and family events. Therefore, SLID opens new research avenues that will provide greater insights on important issues, such as how many Canadians remain in low income situations and what makes it possible for others to emerge from periods of low income.

The comprehensive data that make SLID so valuable, also makes it more complex for Statistics Canada to ensure that confidentiality of respondents is maintained.

In order to comply with the strict confidentiality provisions of the Statistics Act, SLID longitudinal data are made available through new modes of dissemination, namely:

- remote access: computer program(s) are written by clients and sent electronically to Statistics Canada, where staff run the program(s) against the data base and apply confidentiality protection measures. If need be, data are suppressed from the output. Survey officers subsequently return results to clients;

- on premise access: researchers under contract with Statistics Canada are given access to Regional Reference Centres across the country, where staff provides data retrieval infrastructure and implements confidentiality procedures;

- research data centres: were opened in 2000 on selected university campuses across the country. These centres will act as extensions of Statistics Canada and provide researchers with access to the data, while protecting confidentiality.

Research and working papers

Statistics Canada publishes a variety of research and working papers that are made available free of charge on its website (www.statcan.ca). Listed below is a selection of recent papers for readers interested in income trends. Several other reports are also available.

- Effects of Self-Rated Disability and Subjective Health on Job Separation 750002MIE2002001
- Recent Developments in the Low Income Cutoffs 750002MIE2001003
- Should the Low Income Cutoffs be Updated? A Summary of Feedback on Statistics Canada's Discussion Paper 75F0002MIE2000011
- To What Extent are Canadians Exposed to Low Income? 75F0002MIE1999001
- The Persistent Gap: New Evidence on the Canadian Gender Wage Gap 75F0002MIE1999008
- A Comparison of the Results of the Survey of Labour and Income Dynamics (SLID) and the Survey of Consumer Finances (SCF) 1993-1997: Update 75F002MIE1999007
- Low-income Intensity during the 1990s: the Role of Economic Growth, Employment Earnings and Social Transfers 11F0019MIE2003172
- Wage Progression of less Skilled Workers in Canada: Evidence from the SLID (1993-1998) 11F0019MIE2002194
- Wives, Mothers and Wages: Does Timing Matter? 11F0019MIE2002186
- The Performance of the 1990s Canadian Labour Market 11F0019MIE2000148

The menu path to download the papers listed above is "Our Products and Services" then "Browse our Internet Publications (free)" followed by the catalogue number.

SLID documentation for researchers

- Survey Overview – Survey of Labour and Income Dynamics 75F0011XIE
- Survey of Labour and Income Dynamics Microdata User's Guide 75M0001GIE
- SLID Electronic Data Dictionary 75F0026XIB
- SLID questionnaires 75F0002MIE1999003, 75F0002MIE1999004, 75F0002MIE1999005

The menu path to download the papers listed above is "Our Products and Services" then "Browse our Internet Publications (free)" followed by the catalogue number.

Perspectives on Labour and Income 75-001-XPE

Perspectives on Labour and Income is a quarterly journal that features analytical articles on the latest trends. It includes a section that summarizes recent reports and studies released by Statistics Canada. Subscribing to Perspectives on Labour and Income will prove to be an excellent way to keep up-to-date on what's new, all year long!

11. Questions and comments

If you have any questions or comments about the data in this CD-ROM product, you can contact the Income Statistics Division.

Telephone: 1-888-297-7355 or 613-951-7355
Facsimile Number: 613-951-3012
Internet: income@statcan.ca

Income Statistics Division
Statistics Canada
Ottawa, Ontario
K1A 0T6